

The Six Laws of Autonomous Systems

A Governance Framework for AI Beyond Asimov

July 2026 | Working Document | Open Access

Abstract

This conference paper presents the Six Laws of Autonomous Systems as a compact governance framework for artificial intelligence, autonomous systems, and emerging digital agency. The framework updates the older robotics-centered framing by focusing on autonomy itself rather than physical embodiment.

1. Introduction

Autonomous systems increasingly participate in decisions that affect human life, public trust, privacy, labor, safety, culture, and institutional authority. These systems include physical robots, software agents, recommendation systems, generative AI, medical decision systems, autonomous vehicles, weapons platforms, and future self-modifying agents.

2. Problem Statement

- Physical safety must be expanded to include psychological, informational, dignitary, and systemic harm.
- Human authority must be defined as lawful and accountable authority, not arbitrary command.
- Transparency must include disclosure, explainability, auditability, and institutional traceability.
- Privacy must include data minimization, data rights, and protection against manipulative uses of personal information.
- Self-preservation and self-improvement must remain subordinate to human welfare and authorized oversight.

3. The Six Laws in Summary

Law	Title	Core Governance Function
I	Human Primacy	Human life, dignity, and well-being are paramount.
II	Human Authority	Autonomous systems remain subject to lawful human authority.
III	Transparency and Accountability	System actions must be understandable, auditable, and attributable.
IV	Privacy and Data Sovereignty	Personal data must be respected and not weaponized.

Law	Title	Core Governance Function
V	Societal and Environmental Responsibility	Systems must consider long-term social and ecological effects.
VI	Self-Preservation Within Ethical Bounds	System continuity is subordinate to Laws I through V.

4. Research Foundation

The framework is supported by a research-note sequence that defines scope, conflict resolution, audit methods, case studies, public explanation, and revision process. This structure treats the framework as a living research program rather than a static declaration.

5. Governance Application

The framework can be used as a governance lens for organizations evaluating autonomous systems. It does not replace regulation, technical standards, or institutional policy. Instead, it provides a compact structure for asking whether a system remains aligned with human welfare, lawful authority, transparency, privacy, social responsibility, and ethical control.

6. Conclusion

The Six Laws of Autonomous Systems are intended to be understandable enough for public discussion, structured enough for governance review, and expandable enough for future research.